Analysis of Vaguely Classified Data

by

Kwok-leung HO

A

Thesis

submitted to

(Division of Statistics)

the Graduate School

of

The Chinese University of Hong Kong

In Partial Fulfillment

of the Requirements for the Degree of

Master of Philosophy

(M Phil.)

June, 1994

fr. Li Kin Hung

UL

Thesis
QA
278
H58
1994

14 SEP 1995

THE CHINESE UNIVERSITY OF HONG KONG

GRADUATE SCHOOL

The undersigned certify that we have read a thesis, entitled "Analysis of Vaguely Classified Data" submitted to the Graduate School by Ho Kwok Leung (何 國 良) in partial fulfillment of the requirements for the degree of Master of Philosophy in Statistics. We recommend that it be accepted.

Kim-Hung Li

Dr. K.H. Li
Supervisor

Chan Ping Shing

Dr. P.S. Chan

Leung Pui Lam

Dr. P.L. Leung

Prof. K.L. Tsui
External Examiner

DECLARATION

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

ACKNOWLEDGEMENT

I would like to express my sincere gracefulness to my supervisor, Dr. K.H. Li, for his supervision and encouragement during the course of this research program. It is a pleasure to extend my gratitude to all the staff of the Department of Statistics. I would like to thank Miss Betty Fung for her supports.

ABSTRACT

It is common for one to encounter with some vaguely classified variables, like tallness and cleverness. In this thesis, some basic beliefs on the vaguely classified variable are introduced. Based on these beliefs, a random threshold model, instead of a fixed threshold model, is introduced to handle the vaguely classified variables which are observed in the form of contingency table. A fundamental case of the model is used to demonstrate the estimation part of the model. The method adopted in the estimation is Gibbs sampler.

Key words : Contingency table, Gibbs sampler, Threshold, Vaguely classified variable.

CONTENTS

Chapter 1 Introduction

We always come across some qualitative variables. For example, we usually use the words like 'Very bad', 'Bad', 'Fair', 'Good', 'Very good', etc. We always classify people as 'Tall' or 'Short'. However, it is very often you meet the case that someone says, ' Hey! Look! How tall that guy is!' but you do not agree with him after you have seen that guy. Why does it happen? It is because it is unclear whether a specific person should be described as 'tall'. There is no precise height that separates tall from 'non-tall' persons. Those categorical variables with each category vaguely defined are referred to as vaguely classified variables. Categorical variable taking values 'tall' and 'short' is obviously vaguely classified variable. In the following study, we are dealing with this kind of variables.

There is a parallel aspect in the set theory. In classical set theory, an element can either belong to a set or not. Thus a classical set S_{160}, the set of people with height greater than 160cm, has no vagueness. Every element in the population can be classified to be in S_{160} or not in S_{160}. However, instead of S_{160}, we may want to consider the set of tall people, S_{tall}. In this situation, the classical set theory fails. Zadeh (1965) introduced the theory of fuzzy sets which aimed at modelling this vagueness. Instead of classifying a person x as 'tall' or 'not tall', a degree of membership, $\mu_{S_{tall}}(x)$, between zero and one will be assigned to every element x in the universal set. The person who is definitely tall has $\mu_{S_{tall}}(x)=1$, one who is definitely not tall has $\mu_{S_{tall}}(x)=0$, and intermediate values are used for the ambiguous cases.

Although we clearly know the mathematical definition of a fuzzy set, there is still ambiguity in the theory of fuzzy sets. An obvious example is the interpretation of the membership value. There are relatively few published papers which focus on the interpretation of membership values. Among those limited number of papers, there are several interpretations of membership values (Osherson and Smith (1981, 1982), Giles (1982, 1988), Norwich and Turksen (1984), Hisdal (1986a, 1986b, 1988a, 1988b), Smithson (1986), Turksen (1991) and Mabuchi (1992)). A unique, general and convincing solution on the interpretation of membership values has not been proposed. For example, according to Giles, the membership value of element x in the fuzzy set A, $\mu_A(x)$ is that the subject agrees that the following is a "fair bet":

> 'If you pay me \$ $\mu_A(x)$ then I will agree to pay \$ 1 when the test-procedure for A gives outcome "yes", given a test-procedure for accessing pass or failure of x is defined.'

It appears that the assignment of $\mu_A(x)$ becomes simply a special case of assignment of a subjective probability. However, there are still some defects in this definition. First of all, it assumes that for any fuzzy set, there is a corresponding test. However, it is not always true. For example, if we concern a set of perfect man. It is hard to find a test for examining whether the person is perfect or not. Secondly, even when you may have a series of tests to examine this fuzzy concept, the test you have may not be the same as that of mine. Thirdly, Chow (1993) stated that

> 'Under this interpretation $A \cap A \neq A$ for many fuzzy set A. For example, when the set is the set of 'unbreakable glass' and the test procedure is that drop the glass from a height of 5 feet onto a wooden floor; if it breaks, the outcome is "no"; otherwise the outcome is "yes". Then when we consider $A \cap A$, we need to do the test twice and we assign a positive outcome only if the glass breaks on neither test. Obviously, it is equivalent to a harder test than that of A. In other words, $A \cap A \neq A$ under this interpretation.'

Therefore, it seems that the fuzziness comes from the definition of the test. If a test is defined and everyone knows the test then it becomes a problem of subjective probability. In this thesis, instead of dealing with the fuzziness of membership functions, a different approach that makes use of thresholds will be adopted.

In order to analyze vaguely classified data, we assume that for each vaguely classified variable, there exists a latent variable, which is a discrete or continuous variable. Classification bases on the value of the latent variable and some values called thresholds. Let us see how thresholds work in the following example.

Example 1.1

Consider a vaguely classified variable which is the classification of students made by their teacher. Suppose there are five grades for the students; they are 'Very bad', 'Bad', 'Fair', 'Good' and 'Very good'. Moreover, the latent variable is the examination score of the students. Now, suppose that the thresholds used by the teacher in classifying the student are 20, 40, 70 and 90 respectively. If the examination mark of a student is denoted by X, then

$$\text{Comment} = \begin{cases} \text{Very Bad} & \text{if} \quad X \le 20 \\ \text{Bad} & \text{if} \quad 20 < X \le 40 \\ \text{Fair} & \text{if} \quad 40 < X \le 70 \\ \text{Good} & \text{if} \quad 70 < X \le 90 \\ \text{Very Good} & \text{if} \quad 90 < X \end{cases}$$

Example 1.1 is a typical example for the analysis making use of thresholds. Recently, threshold model is widely adopted in different fields of study in statistical analysis. For example, in marketing research, threshold model is used to investigate the effects of question form, question content, and respondent experience/involvement factors on indicator of "don't know" item nonresponse to attitude questions (Leigh and Martin 1987). DeSarbo and Cho (1989) used threshold model to analyze "pick any/n" choice data (e.g. consumers rendering buy/no buy decisions concerning a number of actual products). For behavior research, Helmes and Jackson (1989) used threshold model to analyze the personality item responding. In time series analysis, Tong (1990) discussed the threshold model in the non-linear time series analysis. Olsson (1979) used threshold model in estimating polychoric correlation coefficient.

Thresholds provide a simple way to handle the vaguely classified variable. However , the classification depends on the participant. As illustrated in the previous example, if you got 90 marks in your examination and your comment of your teacher was 'Good', you may argue," I have tried my best in my examination, I should worth a comment of 'Very Good'. How cruel is the teacher!" Such disagreement is common in daily life. Why does it happen? The answer is simple. As every person has his point of view, the thresholds are clearly individual dependent.

Nowadays practitioners usually assume that the thresholds are fixed but unknown parameters rather than variables. As you can see from the above example, it is not the case in the real life. If fixed threshold assumption is employed, odd result may be obtained as illustrated in following example.

Example 1.2

Given a subject of X cm high, N persons are asked to classify the subject as 'Tall', 'Fair' or 'Short'. If constant thresholds are assumed, same answer is expected. The result must be one of the following situations under the constant thresholds assumption.

Tall	Fair	Short
N	0	0

Tall	Fair	Short
0	N	0

Tall	Fair	Short
0	0	N

It certainly does not agree with what we actually observe in real life. The assumption of constant thresholds seems not justified.

The plan of this thesis is as follows. In chapter 2, section 1 will give the basic beliefs on vaguely classified variables, like the existence of super-population, the formation of the reference set and the aspect of ‘observing with error’. Some properties of the contingency table will be introduced in section 2. In section 3, a mathematical model is proposed. The general one will be formulated first. Then variations of the mathematical model will be discussed for different special cases. In chapter 3, a study of a special case of the proposed model will be shown. A basic model of dimension two will be used to illustrate the ideas discussed in the previous parts. Detailed simulation methods will be presented. Finally, chapter 4 summarizes our discussion and suggests further studies and investigations.

Chapter 2 Model

2.1 Basic beliefs on the vaguely classified variable

As discussed in chapter 1, we see that people seldom use same set of thresholds to classify the subjects. Therefore, the formation of thresholds for the individuals is a crucial part in dealing with the vaguely classified variable. In order to understand the idea, let us examine the following example.

Example 2.1.1

Suppose subjects are classified into five categories which are 'Very clever', 'Clever', 'Fair', 'Unwise' and 'Very unwise'. For simplicity, we denote them as {c1,c2,c3,c4,c5} respectively. Moreover, suppose that the IQ (Intelligence Quotient) is the latent variable for the variable of cleverness. Now, consider the following subjects:-

(1) A genius (IQ=170, say)

(2) The author (IQ=90, say)

(3) An idiot (IQ=40, say)

(4) An extra-terrestrial (IQ=30, say)

How should you classify them into one of the five categories? Before you perform the classification, what else should you know in advance?

First of all, you may want to know what population do you concern. For example, if the interesting population is all the people in the world, you may classify the genius to be category c1 (i.e. Very clever), the author to be category c3 and the idiot and the extra-terrestrial to be category c5. However, if the interesting population is all the people in the University, you may still classify the genius to be category c1 but classify the author to be category c4 and so on. Therefore, the classification depends on the interesting population. To be more extreme, if we consider the population of all living creatures in the Universe (assume we have knowledge about that), then the extra-terrestrial mentioned above may be very clever in that particular interesting population because there may be a lot of living creatures having low IQ marks.

Example 2.1.2

In newspapers, you very often read some reports like the following one:

> 'Team A have been performing very well in this quarter, it can be regarded as a very strong team here; however, when compared with the teams in the world, she has to work harder in order to become a strong team in the world.'

We can see that if different populations are of interest, different classifications are made, i.e. the set of thresholds may vary.

Secondly, besides the interesting population, thresholds depend on the individuals. This idea have been mentioned in chapter 1. Let us continue Example 2.1.1 and consider the following situation in order to understand how thresholds depend on the participants.

Example 2.1.3 (Continued from Example 2.1.1)

Suppose the following individuals are required to classify the subjects mentioned above into the five categories, the resultant classification would probably be as follows:-

Person who classifies \ Subject	The genius	The author	The idiot	The E.T.
Undergraduate	c1	c3	c5	c5
Lecturer	c1	c4	c5	c5
Child	c1	c1	c1	c1

It is obvious that the classifications made by different people are quite different. The formation for this aspect can be explained by the dependence of thresholds on individuals. Since the observations of each individual in the population may be quite different, even they concern the same population, the classification may be different. For example, the intelligence of the set of people that the child knows can be different from that of the undergraduate or the lecturer. Similar situation also exists between the undergraduate and the lecturer. The standards they used to classify the subjects are not the same, and neither do the classifications. In our context, it means that they used different sets of thresholds to classify.

In brief, the thresholds do not only depend on the interesting population but also the sample which the participants observe in the interesting population. We conclude these ideas as the following belief.

Belief 2.1.4

When concerning a vaguely classified variable Z (k categories), a latent variable X is assumed. A super-population exists which characterizes the distribution of X. For each individual i, he uses the set of thresholds $\{\alpha_1^i,\cdots,\alpha_{k-1}^i\}$ (with index i to indicate the thresholds depends on i) to classify the subject into one of k categories by comparing the value of the latent variable of the subject and the thresholds. Moreover, from the arguments above, a sample R_i (called reference set) will be drawn from the super-population which in turn affects the values of thresholds $\{\alpha_1^i,\cdots,\alpha_{k-1}^i\}$.

Following the idea of the Belief 2.1.4, the relationship between the reference set, R_i and the thresholds is concerned. First of all, it is easy to understand that the thresholds have an increasing property, i.e.

$$\alpha_1^i \le \alpha_2^i \le \cdot \cdot \cdot \le \alpha_{k-2}^i \le \alpha_{k-1}^i \ .$$

Therefore, it is quite straightforward and reasonable to believe that the thresholds relate to the order statistics of R_i. The following example illustrates this aspect.

Example 2.1.5

In most of the letters of recommendation, your recommender is required to assess your different kinds of ability by using the table like the following one:

	Excellent (Top 2%)	Good (Top 15%)	Fair (Top 30%)
Academic perfomance	________	________	________
Intellectual potential	________	________	________
Creativity and originality	________	________	________

The values given in the brackets are the criteria for classification. They correspond to the sample percentiles of the reference set for the recommender, which actually are the order statistics of the reference set.

Finally, one aspect that has to be mentioned in this problem is that latent variable may be observed with error. In order to illustrate how this aspect will relate to the present situation and affect one's classification, let us revisit our example again.

Example 2.1.6 (Continued from Example 2.1.3)

Let us consider the situation such that IQ marks of the interested subjects are not given, and the three individuals, the undergraduate, the lecturer and the child, are asked for classifying the subjects. Then, the situation will be different from those in Example 2.1.3 due to the reason of observing with error. The following table summarizes the situation and the decisions of the three individuals.

	The genius	The author	The idiot	The E.T.
	[170]	[90]	[40]	[30]
Undergraduate	(160)	(80)	(50)	(60)
	c1	c4	c5	c4
Lecturer	(170)	(70)	(45)	(35)
	c1	c4	c5	c5
Child	(120)	(25)	(20)	(60)
	c1	c2	c3	c1

Remarks : [True IQ marks]

: (Observed IQ marks)

As you can see from the table, the situation is different from the case that exact measurements are given. In Fuzzy set, a similar argument is found. Hisdal (1986a, 1986b, 1988a, 1988b) listed different sources of fuzziness. The first source of fuzziness is due the individual's recognition that under not exact conditions of observations. An individual may make errors in the estimation of the attribute values of the subjects.

In our present situation, the aspect of 'observing with error' also exists in the formation of the reference set R_i. It is because when a sample is drawn from the super-population, the exact measurements of the elements in the reference set are not necessary known. (You might not know the height of your best friend.) In brief, the aspect of 'observing with error' need to be included in the model of vaguely classified data.

2.2 Properties of the contingency table

In this section, we will discuss some properties of the contingency table. First of all, we define the component-wise strictly increasing transformation, which is important in the following context.

Definition 2.2.1

Component-wise strictly increasing transformation g of vector X = $(X_1, X_2, \ldots, X_p)'$ is defined as

$$\begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix} \xrightarrow{g} \begin{pmatrix} g_1(X_1) \\ \vdots \\ g_p(X_p) \end{pmatrix},$$

where g_i is strictly increasing function of X_i, for $i=1,\ldots,p$.

Recalling the previous discussions, one classifies a subject with observed univariate value x into category j, for $j=1,\cdots,k$ (k categories) if

$$\alpha_{j-1} < x \le \alpha_j ,$$

where α_j is the j th threshold with $\alpha_0 = -\infty$ and $\alpha_k = \infty$. When a strictly increasing transformation h is applied to both the thresholds and observed value of the subject, the inequality still holds in the following sense:-

$$h(\alpha_{j-1}) < h(x) \le h(\alpha_j) .$$

That means the classification of the individual is still category j without any change. As a result, we can conclude that any component-wise strictly increasing transformation on the latent variable and the corresponding thresholds will give the same contingency table. Therefore, based on the observed frequencies of the contingency table, we do not have any information about the mean and the dispersion of the latent variables.

Afterwards, we define a family of distribution in which one distribution is a component-wise strictly increasing transformation of the another. Then all member in the family will fit the observed data equally well provided that prior distributions change accordingly. The family of distribution is closed under component-wise strictly increasing transformation. Thus, we can choose a representative distribution to represent the family of distribution.

A suggestion for the representative distribution is normal distribution with mean vector 0 and unity variance. It is because there is no other distribution in the family except this distribution whose marginal distribution of each component is standard normal distributed and has unity variance. The component-wise strictly increasing transformation of normal distribution with mean vector 0 and unity variance to a distribution in the same family with marginal standard normal distribution of each component must be the identity transformation.

Presently, you may notice that the correlation of $\{X_1,X_2\}$ and that of $\{g_1(X_1),g_2(X_2)\}$ are not the same, where g_1 and g_2 are strictly increasing functions; however, with the respective transformation of thresholds, it gets the same contingency table. Therefore, it is necessary to find the parameter which is invariant to such transformation. Two parameters are proposed:-

(i) corr $(F_{x_1}(X_1),F_{x_2}(X_2))$, denoted by ω_{X_1,X_2}.

(ii) Kendall's Tau, denoted by τ_{X_1,X_2}.

Both parameters are invariant to the component-wise strictly increasing transformation. We would like to find the relationships of all these parameters and the correlation coefficient, ρ. First of all, we give the following lemma, which was proved by Cramér (1946), who used Hermite polynomials and infinite series in his proof. In Kepner, Harper and Keith (1989), same result is proved by a simpler method. Moreover, Stigler (1989) and Farebrother (1989) gave comment on the same bivariate normal orthant probabilities with a simple analytical proof and a geometrical proof.

<u>Lemma 2.2.2 (Cramér (1946))</u>

Let $\Pr(X \geq 0 , Y \geq 0) = \alpha$ for the distribution

$$\begin{pmatrix} X \\ Y \end{pmatrix} \overset{D}{=} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right].$$

Then $\rho = \sin(2\alpha-\frac{1}{2})\pi$ or equivalently $\alpha = \frac{1}{4} + \frac{1}{2\pi}\sin^{-1}\rho$.

Theorem 2.2.3 (Kepner, Harper and Keith (1989))

If

$$\begin{pmatrix} X \\ Y \end{pmatrix} \overset{D}{=} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right],$$

then Kendall $\tau_{X,Y} = \frac{2}{\pi}\sin^{-1}\rho$.

Theorem 2.2.4

If

$$\begin{pmatrix} X \\ Y \end{pmatrix} \overset{D}{=} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right],$$

then

$$\omega_{X,Y} = \mathrm{Corr}(\Phi(X),\Phi(Y)) = \frac{6}{\pi}\sin^{-1}(\frac{\rho}{2})$$

Proof

By working with the definition of expectation, we have

$$E(\Phi(X)\Phi(Y)) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{x}\phi(u)\,du\right]\left[\int_{-\infty}^{y}\phi(v)\,dv\right] f(x,y)\,dx\,dy,$$

where $\phi(\cdot)$ is the pdf of standard normal variable and $f(x,y)$ is the joint pdf of X and Y. Consider the probability, Pr $(U \le X, V \le Y)$ where U, V are standard normal distributed and $(X,Y)'$ is distributed as stated above. Moreover, U, V and $(X,Y)'$ are independent. Then,

$$\Pr(U \le X,\ V \le Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{x} \phi(u)\,du \int_{-\infty}^{y} \phi(v)\,dv\, f(x,y)\,dx\,dy .$$

Therefore,

$$E(\Phi(X)\Phi(Y)) = \Pr(U-X \le 0,\ V-Y \le 0) .$$

Note that

$$\begin{pmatrix} U-X \\ V-Y \end{pmatrix} \overset{D}{=} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & \rho \\ \rho & 2 \end{pmatrix}\right],$$

By lemma 2.2.2, we have

$$E(\Phi(X)\Phi(Y)) = \frac{1}{4} + \frac{1}{2\pi}\sin^{-1}(\frac{\rho}{2}) .$$

Then, consider

$$\begin{aligned} \mathrm{Corr}(\Phi(X),\Phi(Y)) &= \frac{1}{\frac{1}{12}}\mathrm{Cov}(\Phi(X),\Phi(Y)) \\ &= 12\ \mathrm{Cov}(\Phi(X),\Phi(Y)) \\ &= 12\left(E(\Phi(X)\Phi(Y)) - 1/4\right), \end{aligned}$$

because

$$E(\Phi(X)) = E(\Phi(Y)) = \frac{1}{2},$$

$$\mathrm{Var}(\Phi(X)) = \mathrm{Var}(\Phi(Y)) = \frac{1}{12},$$

as $\Phi(X)$ and $\Phi(Y)$ are both distributed as $U(0,1)$. Therefore,

$$\begin{aligned} \omega_{X,Y} &= \mathrm{Corr}(\Phi(X),\Phi(Y)) = 12 \times (\frac{1}{4} + \frac{1}{2\pi}\sin^{-1}(\frac{\rho}{2}) - \frac{1}{4}) \\ &= \frac{6}{\pi}\sin^{-1}(\frac{\rho}{2}) \end{aligned}$$

Q.E.D.

Theorem 2.2.5

If

$$\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \begin{pmatrix} X_2 \\ Y_2 \end{pmatrix}, \ldots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \overset{iid}{\sim} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right],$$

then

$$\text{Corr}\ (R_n(X_1), R_n(Y_1)) = \left(\frac{12}{n+1}\right)\left(\frac{1}{2\pi}\sin^{-1}\rho + (n-2)\frac{1}{2\pi}\sin^{-1}(\tfrac{\rho}{2})\right),$$

where $R_n(X)$ is the rank of X in a sample of size n.

Proof

By the definition of correlation,

$$\text{Corr}\ (R_n(X_1), R_n(Y_1)) = \frac{E(R_n(X_1)R_n(Y_1)) - E(R_n(X_1))E(R_n(Y_1))}{\sqrt{\text{Var}(R_n(X_1))\text{Var}(R_n(Y_1))}}.$$

Note that

$$R_n(X_1) \overset{D}{=} \text{Discrete Uniform}\ (1,n),$$

$$E(R_n(X_1)) = (n+1)/2,$$

$$\text{Var}(R_n(X_1)) = (n^2-1)/12,$$

and the situation is the same for that of Y_1. Consider

$$E(R_n(X_1)R_n(Y_1))$$

$$= E\left[(1 + \sum_{i=2}^{n} I_{X_i<X_1})(1 + \sum_{j=2}^{n} I_{Y_j<Y_1})\right],$$

where I_A is the indicator function such that

$$I_A = \begin{cases} 1 \text{ if } A \text{ is true} \\ 0 \text{ if otherwise} \end{cases}.$$

Therefore,

$$E(R_n(X_1)R_n(Y_1))$$

$$= E\left[1+\sum_{i=2}^{n} I_{X_i<X_1}+\sum_{j=2}^{n} I_{Y_j<Y_1}+\sum_{i=2}^{n} I_{X_i<X_1} I_{Y_i<Y_1}\right.$$

$$\left.+\sum_{i\neq j=2}^{n} I_{X_i<X_1} I_{Y_j<Y_1}\right]$$

$$= 1+\sum_{i=2}^{n} E(I_{X_i<X_1})+\sum_{j=2}^{n} E(I_{Y_j<Y_1})+\sum_{i=2}^{n} E(I_{X_i<X_1} I_{Y_i<Y_1})$$

$$+\sum_{i\neq j=2}^{n} E(I_{X_i<X_1} I_{Y_j<Y_1})$$

$$= 1+\frac{1}{2}(n-1)+\frac{1}{2}(n-1)+(n-1)\Pr(X_i<X, Y_i<Y_1)$$

$$+ (n-1)(n-2)E(E(I_{X_i<X_1} I_{Y_j<Y_1} | X_1, Y_1)) ,$$

for any $i,j = 2,\cdots,n$ and $i \neq j$, because $\Pr(X_i<X_1)=1/2$, as $X_{i,} X_1$ are iid for any $i = 2,\cdots,n$. Therefore,

$$E(R_n(X_1)R_n(Y_1))$$

$$= n + (n-1)\Pr(X_i<X_1, Y_i<Y_1) + (n-1)(n-2)E(\Phi(X_1)\Phi(Y_1))$$

because

$$E\Big(\Pr(X_i<X_1, Y_j<Y_1 | X_1, Y_1)\Big) = E(\Phi(X_1), \Phi(Y_1)) ,$$

as X_i, Y_j are iid if $i\neq j$. Note that,

$$\begin{pmatrix} X_i-X_1 \\ Y_i-Y_1 \end{pmatrix} \overset{D}{=} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, 2\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right] .$$

By lemma 2.2.2,

$$\Pr(X_i-X_1<0, Y_i-Y_1<0) = \frac{1}{4} + \frac{1}{2\pi}\sin^{-1}\rho .$$

Moreover, by the proof of theorem 2.2.4, we know

$$E(\Phi(X_1)\Phi(Y_1)) = \frac{1}{4} + \frac{1}{2\pi}\sin^{-1}(\frac{\rho}{2}) .$$

Therefore,

$$\mathrm{Corr}\ (R_n(X_1),R_n(Y_1))$$

$$= \frac{n+(n-1)(\frac{1}{4} + \frac{1}{2\pi}\sin^{-1}\rho) + (n-1)(n-2)(\frac{1}{4} + \frac{1}{2\pi}\sin^{-1}(\frac{\rho}{2})) - (\frac{n+1}{2})^2}{\frac{n^2-1}{12}}$$

$$= \left(\frac{12}{n+1}\right)\left(\frac{1}{2\pi}\sin^{-1}\rho + (n-2)\frac{1}{2\pi}\sin^{-1}(\frac{\rho}{2})\right) . \qquad \text{Q.E.D.}$$

<u>Corollary 2.2.6</u>

For the setting of Theorem 2.2.5,

$$\lim_{n\to\infty} \mathrm{Corr}\ (R_n(X_1),R_n(Y_1)) = \mathrm{Corr}\ (\Phi(X_1),\Phi(Y_1)) = \omega_{X_1,Y_1} .$$

Proof

By theorem 2.2.5

$$\mathrm{Corr}\ (R_n(X_1),R_n(Y_1)) = \left(\frac{12}{n+1}\right)\left(\frac{1}{2\pi}\sin^{-1}\rho + (n-2)\frac{1}{2\pi}\sin^{-1}(\frac{\rho}{2})\right) .$$

Therefore,

$$\lim_{n\to\infty} \mathrm{Corr}\ (R_n(X_1),R_n(Y_1))$$

$$= \lim_{n\to\infty} \left(\frac{12}{n+1}\right)\left(\frac{1}{2\pi}\sin^{-1}\rho + (n-2)\frac{1}{2\pi}\sin^{-1}(\frac{\rho}{2})\right)$$

$$= \frac{6}{\pi}\sin^{-1}(\frac{\rho}{2})$$

$$= \text{Corr}\,(\Phi(X_1),\Phi(Y_1))$$

$$= \omega_{X_1,Y_1} \,. \qquad\qquad \text{Q.E.D.}$$

Remarks :

(i) 'No tie' is assumed. It is reasonable for (X,Y) is continuous.

(ii) If

$$\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \begin{pmatrix} X_2 \\ Y_2 \end{pmatrix}, \ldots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \overset{iid}{=} F \,,$$

where F is an arbitrary continuous distribution, then

$$\lim_{n\to\infty} \text{Corr}\,(R_n(X_1),R_n(Y_1)) = \text{Corr}\,(F(X_1),F(Y_1)) = \omega_{X_1,Y_1} \,.$$

(iii) By Corollary 2.2.6, we can see that the interesting parameter, ω_{X_1,Y_1} is actually the limiting case of the Spearman rank correlation coefficient when n tends to infinity. Therefore, ω_{X_1,Y_1} is a reasonable parameter for the present situation as the size of the super-population is assumed to be infinite.

2.3 Mathematical Model

In this section, we define a mathematical model of the vaguely classified data. First of all, we state the situation we faced. Then according to the situation, we introduce the notations and assumptions which help us to define the mathematical model. Finally, the mathematical model is defined.

In the rest of this section, a special case of the model is included in order to illustrate how to make use of the proposed model to deal with the situation mentioned.

2.3.1 Situation

N individuals are questioned about d subjects. For each subject, p characteristics are concerned. Individual uses his own set of thresholds, and compares them with the latent value of the subject to classify the subject into certain category. Finally, d·p way contingency table is obtained.

In the present situation, the vaguely classified variable is a p-dimensional vector, For each individual i, the classification of subject t according to the p variables is denoted as

$$Z_{it} = \begin{pmatrix} Z_{it1} \\ Z_{it2} \\ \vdots \\ Z_{itp} \end{pmatrix} ,$$

where $i=1,\cdots,N$ (number of individuals); $t=1,\cdots,d$ (number of subjects). Suppose there are $n(v)$ categories for the v th vaguely classified variable, for $v = 1,\cdots,p$. i.e.

$$Z_{itv} = 1,\cdots,n(v) \ , \ v=1,\cdots,p.$$

2.3.2 Assumptions

(1) For the vaguely classified variable Z, we assume that there exists a continuous latent variable Y, which is also an

p-dimensional vector. Moreover, we assume that there exists a super-population, R, of the latent variable Y.

(2) For individual i, a reference set R_i is formed which is the set of the latent variable values of all elements that the individual ever meets. The size of the reference set is denoted by m(i). Therefore, for i=1,···,N,

$$R_i = \left\{ Y_{ij} \text{ , for } j = 1,\ldots,m(i) \right\},$$

where Y_{ij} is the latent variable such that

$$Y_{ij} = \begin{pmatrix} Y_{ij1} \\ Y_{ij2} \\ \vdots \\ Y_{ijp} \end{pmatrix}.$$

(3) We assume that the thresholds are the order statistics of reference set of each individual. Denote the order number of the thresholds by $r_{v,1},\cdots,r_{v,n(v)-1}$, for v = 1,···,p. We require $r_{v,e}$ depends on m(i) in such a way that

$$\phi(v,e,i) = \frac{r_{v,e}}{m(i)}$$

converges in distribution to a fixed value for v=1,···,p; e=1,···,n(v)-1. This requirement emphasizes that the order numbers are to allocate the relative location in the reference set, not the absolute location.

We denote the thresholds of individual i to be $\xi_{i,v,1}$, $\xi_{i,v,2},\ldots,$ $\xi_{i,v,n(v)-1}$ for the v th variable, v=1,···,p, where

$\xi_{i,v,e}$ represents the $r_{v,e}$ th order statistics of v th component of the latent variable in R_i. Moreover, it is also the e th threshold for the vaguely classified variable Z_{itp}. Denote the observed value of the t th subject for the individual i to be

$$W_{it} = \begin{pmatrix} W_{it1} \\ W_{it2} \\ \vdots \\ W_{itp} \end{pmatrix}.$$

Then, $Z_{it} = \begin{pmatrix} Z_{it1} \\ Z_{it2} \\ \vdots \\ Z_{itp} \end{pmatrix}$ if the following conditions fulfill,

$$\xi_{i,v,Z_{itv}-1} < W_{itv} \le \xi_{i,v,Z_{itv}},$$

where $\xi_{i,v,0} \equiv -\infty$ and $\xi_{i,v,n(v)} \equiv \infty$, for $t = 1,\cdots,d$; $v = 1,\cdots,p$; $i = 1,\cdots,N$.

(4) 'Observe with error' exists: As described in the previous section, we may not observe the true value of the latent variables of the interesting subjects. Moreover, when forming the reference set, we also may not observe the true value of the latent variables of the elements in the reference set.

2.3.3 Mathematical Model

In this section, we define a mathematical model of the situation. For individual i, elements in R_i are denoted by Y_{ij} where

$$Y_{ij} = \begin{pmatrix} Y_{ij1} \\ Y_{ij2} \\ \vdots \\ Y_{ijp} \end{pmatrix} \text{ i.i.d. } N(\mu_Y, \Sigma_Y + \Sigma_W) \text{ , for } i=1,\cdots,N \text{ ,}$$

where μ_Y and Σ_Y are the parameter of the super-population R which contains values i.i.d. distributed as $N(\mu_Y, \Sigma_Y)$ and where Σ_W represents the dispersion of observation due to error. If $\Sigma_W = 0$, the problem reduces to the case of observing without error. For subject t, each individual i observes $W_{it} = (W_{it1}, W_{it2}, \cdots, W_{itp})'$ such that

$$W_{it} | T_t \text{ i.i.d. } N(T_t, \Sigma_W) \text{ ,}$$

where T_t is the true value of the subject t, for $t = 1,\cdots,d$.

In the mathematical model, we have the following parameters:-

(1) Distribution parameters : $\mu_Y, \Sigma_Y, \Sigma_W$;

(2) Thresholds parameters : $r_{v,1}, \cdots, r_{v,n(v)-1}$, for $v = 1,\cdots,p$;

(3) Reference set parameters : m(i), for $i = 1,\cdots,N$;

(4) Subject parameters : T_v, for $v=1,\cdots,d$.

For the given observed data, we cannot identify all parameters above. Same contingency table will be obtained by any linear transformation applying on the latent variables and the corresponding thresholds. Based on the observed frequencies of the contingency table, we do not have any information about the mean and the dispersion of the latent variables. In other words, we only have information above the standardized magnitudes of the values of the latent variable and the thresholds.

A common practice to solve the unidentifiable problem is to impose some constraints to the parameters. In the present proposed model, we may either fix the means and variances of the super-population or fix the latent values of the subjects, T and the variances of super-population.

In order to let the reader familiar with the notations introduced and how the situations relate to the model, several examples are introduced.

<u>Example 2.3.3.1</u>

Chow (1993) studied the estimation of membership functions. A survey was conducted. One part of the survey studied the vaguely classified variable, youngness. Individuals are required to classify people in the population of Hong Kong citizens with certain ages to be 'young' and 'not young'. For example, individuals are questioned to classify the subject with age 15 to be 'young' or 'not young'. If more vaguely classified variables, like tallness, fatness are considered, this study will become a case of the proposed model, which can be used to study the relationships between the vaguely classified variables.

Since the exact latent values of the subject are given, if we further assume that the individuals can observe the exact latent values of the elements in the reference set, the situation will be the case of 'multiple individuals, multiple subjects, observing without error'.

Example 2.3.3.2

Benjafield et al (1993) studied the imagery, concreteness, goodness and familiarity ratings for 500 English proverbs. 120 undergraduate student volunteers participated in the study. Each proverb was shown to the volunteers one by one and the volunteers were required to give the ratings for the proverb about these four variables. A 7-point scale, anchored by the words defining the scale (e.g., low imagery vs. high imagery) system was adopted for the undergraduate to rate.

This study is a typical situation discussed in the thesis and can be classified as 'multiple individuals, multiple subjects and observing with error' case. Four vaguely classified variables ($p=4$), imagery, concreteness, goodness and familiarity of English proverbs, were studied. Each of them has 7 categories ($n(v)=7$, $v=1,\cdots,4$). Totally, there are 500 subjects ($d=500$) and 200 individuals ($N=200$). Since the volunteers do not know the latent values of four latent variables of those 500 English proverbs and neither all other English proverbs in the world, it becomes the case of 'observing with error'. It is obvious that everyone knows different number of proverbs, the size of the reference set depends on individual.

There are still many other cases for different setting of the situations. For example it may be the case of 'multiple individuals, single subject, observing with error'. The study of the governor's eloquence and performance is a typical example for this case. In the study, a number of citizens are questioned about these two attitudes of the governor and classify them in certain category. In the following section, a special case of the model is studied. Detailed model specifications and the simulation methods are given in order to illustrate the estimation methods. The reasons of choosing this model are that firstly it is a fundamental model. It is simple enough to demonstrate the model specifications and the simulation methods. Secondly, the modifications of this fundamental model to any other more general model, like multidimensional model, are possible. Finally, the experiences in dealing with this model are useful when facing more complicated setups.

2.3.4 A special case of the model

Let us illustrate how to deal with the above setup through the following example:-

Example 2.3.4.1

Suppose N individuals are questioned about two characteristics (p=2) of one subject (d=1). For example, we are interested in the highness and fatness of the author. The height and weight are the latent variables for the vaguely classified variables respectively. Here, we assume that the case is 'observing without error' i.e. $\Sigma_W=0$. Therefore, for the formation of reference set, the individual can observe the true values of the two latent variables of the elements in the super population. For the size of the reference population, we assume that they are the same and known. That is,

$$m(i) = m \text{ (say)} , \text{ for } i = 1,\cdots,N .$$

Moreover, given m, we assume the order number of the threshold $r_{v,e}$ degenerates, $e=1,\cdots n(v)-1$; $v=1,\cdots,p$.

(1) Model : For individual i, elements in R_i are denoted by Y_{ij} where

$$Y_{ij} = \begin{pmatrix} Y_{ij1} \\ Y_{ij2} \end{pmatrix} \text{ i.i.d. } N(\mu_Y,\Sigma_Y) , \text{ for } j = 1,\cdots,m ,$$

where μ_Y and Σ_Y are the parameters of the super-population R. For the subject, each individual i observes W_{i1} s.t. $W_{i1} = T$ where T is the true value of the subject.

(2) Prior distribution : As we are making use of the Bayesian approach to analyze the problem, we have to give a prior distribution of the parameters in advance. The parameters in the model are μ_Y,Σ_Y, T and the order number of the thresholds, $r_{v,1},\ldots,r_{v,n(v)-1}$, for v = 1,2. As discussed above, the contingency table gives us no information about the means and the variances of the latent variables, we fix the means and variances of the latent variables to some pre-assigned constants.

In this thesis, we fix the mean vector to 0 and the variances to 1. The parameters of interest are the true value of the subject, T, the correlation, ρ, of the two latent variables and the order number of the thresholds, $r_{v,1},\ldots,r_{v,n(v)-1}$, for v = 1,2 and their prior distributions are as follows:-

The prior distributions of ρ, T, $r_{1,1},\ldots,r_{1,n(1)-1}$ and $r_{2,1},\ldots,r_{2,n(2)-1}$ are independent.

(a) $\rho \sim U(-1,1)$.

(b) $T \sim U(-\infty,+\infty)\times(-\infty,+\infty)$.

(c) $r_{1,1},\ldots,r_{1,n(1)-1}$ have the same joint distribution as the order statistics of size n(1)-1 from discrete Uniform(1,m).

(d) $r_{2,1},\ldots,r_{2,n(2)-1}$ have the same joint distribution as the order statistics of size n(2)-1 from discrete Uniform(1,m).

The reasons to choose these distributions as prior distributions of the parameters are that firstly, for the situations of ρ and T, it is suitable and natural to assume the priors to be uniformly distributed when we do not have much knowledge about the behavior of the parameters. Secondly, $r_{v,e}$'s, $v=1,\cdots,p$; $e=1,\cdots,n(v)-1$ are the order number of the thresholds, as they are integers range from 1 to m in non-decreasing order for each v, i.e.

$$r_{v,1} \le r_{v,2} \le \cdots \le r_{v,n(v)-1},$$

where $v=1,\cdots,p$. We do not have much knowledge about $r_{v,e}$'s, so we assume they have the same joint p.d.f of the order statistics of discrete uniform(1,m)

Making use of Gibbs sampler, we can simulate the complete data set and find the estimates of the parameters of interest. In the following chapter, details of the simulation study and the simulation results of this special case will be given.

Chapter 3 Simulation

In the previous chapter, we have defined a general mathematical model. Moreover, we have discussed a special case of the model (Example 2.3.4.1). In this chapter, we will restrict ourselves to this special case, and discuss how to use Gibbs sampler to estimate the interesting parameters.

3.1 Likelihood function for the model and the simulation method

3.1.1 Likelihood function

We continue to use the setup of Example 2.3.4.1 to find the likelihood function of the model and to demonstrate the imputation method. First of all, let us recall the setup of Example 2.3.4.1. For individual i, there is a reference set $R_i = \{Y_{ij} : j=1,\cdots,m\}$, whose the elements Y_{ij}'s are identically and independently distributed as $N(\mu_Y,\Sigma_Y)$. It is supposed that all individuals observe the true value $T=(T_1,T_2)'$ of the subject (only one subject is considered in the example).

As the means and variances of the latent variables are unidentificable in the model, we fix the means to zero and the variances to one. For the remaining unknown parameters, their prior distribution are chosen to be independent such that ρ is distributed as $U(-1,1)$, T distributed as $U(-\infty,+\infty)\times(-\infty,+\infty)$ and $r_{v,1},\ldots,r_{v,n(v)-1}$ have the same joint distribution as the order statistics of size n(v)-1 from discrete Uniform(0,m), for v=1,2.

Let $Y_{ij}=\begin{pmatrix} Y_{ij1} \\ Y_{ij2} \end{pmatrix}$. Define

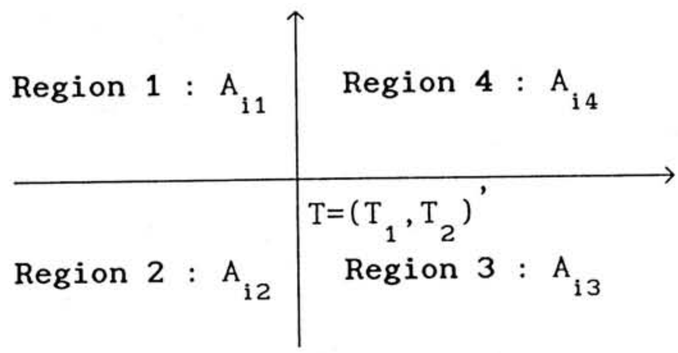
(1) A_{i1} = number of j $(1 \le j \le m)$ such that $Y_{ij1}<T_1$ and $Y_{ij2}>T_2$;

(2) A_{i2} = number of j $(1 \le j \le m)$ such that $Y_{ij1}<T_1$ and $Y_{ij2}<T_2$;

(3) A_{i3} = number of j $(1 \le j \le m)$ such that $Y_{ij1}>T_1$ and $Y_{ij2}<T_2$;

(4) A_{i4} = number of j $(1 \le j \le m)$ such that $Y_{ij1}>T_1$ and $Y_{ij2}>T_2$.

In other words, they denote the number of elements in R_i in the respective region shown below:-



Moreover, we denote P_1,P_2,P_3,P_4 to be the probability on the respective region, i.e.,

$$P_e = \Pr\left((Y_{ij1},Y_{ij2})' \in \textbf{Region e} \right) , \quad e = 1,\cdots,4 .$$

The reason of using A_{ij}'s is that we note the following relationships:-

$$Z_{i1}=z_{i1} \text{ if and only if } r_{1,z_{i1}-1} \le A_{i1}+A_{i2} < r_{1,z_{i1}} ;$$

$$Z_{i2}=z_{i2} \text{ if and only if } r_{2,z_{i2}-1} \le A_{i2}+A_{i3} < r_{2,z_{i2}} , \tag{3.1}$$

where $r_{v,0} \equiv 0$ and $r_{v,n(v)} \equiv m+1$, for $v = 1,2$. Thus, the conditional probability density function for individual i is,

$$f(A_{i1},\ldots, A_{i4}|T,\rho,r_{1,1},\ldots,r_{1,n(1)-1},r_{2,1},\ldots,r_{2,n(2)-1})$$

$$\propto \binom{m}{A_{i1}, A_{i2}, A_{i3}, A_{i4}} P_1^{A_{i1}} P_2^{A_{i2}} P_3^{A_{i3}} P_4^{A_{i4}},$$

with

$$r_{1,z_{i1}-1} \le A_{i1}+A_{i2} < r_{1,z_{i1}},$$

and (3.2)

$$r_{2,z_{i2}-1} \le A_{i2}+A_{i3} < r_{2,z_{i2}},$$

where $r_{v,0} \equiv 0$ and $r_{v,n(v)} \equiv m+1$, for $v = 1,2$. As a result, the joint conditional probability density function for all N individuals becomes,

$$\prod_{i=1}^{N} f(A_{i1}, \ldots, A_{i4} | T, \rho, r_{1,1}, \ldots, r_{1,n(1)-1}, r_{2,1}, \ldots, r_{2,n(2)-1})$$

$$\propto \prod_{i=1}^{N} \binom{m}{A_{i1}, A_{i2}, A_{i3}, A_{i4}} P_1^{A_{i1}} P_2^{A_{i2}} P_3^{A_{i3}} P_4^{A_{i4}},$$

with constraints, $i=1,\ldots,N$

$$r_{1,z_{i1}-1} \le A_{i1}+A_{i2} < r_{1,z_{i1}},$$

and (3.3)

$$r_{2,z_{i2}-1} \le A_{i2}+A_{i3} < r_{2,z_{i2}}.$$

where $r_{v,0} \equiv 0$ and $r_{v,n(v)} \equiv m+1$, for $v = 1,2$. We define the complete data set to be all A_{ij}'s. By relationship (3.1), given the observed frequencies, we obtain inequality (3.3) of A_{ij}'s. The joint conditional probability function of A_{ij}'s can be viewed as the likelihood function of the complete data set. Together with the prior distributions chosen above, the joint posterior distribution of T, ρ, and $r_{1,1},\ldots,r_{1,n(1)-1}$, $r_{2,1},\ldots,r_{2,n(2)-1}$ is proportional to

$$\prod_{i=1}^{N}\binom{m}{A_{i1},A_{i2},A_{i3},A_{i4}} P_1^{A_{i1}}\, P_2^{A_{i2}}\, P_3^{A_{i3}}\, P_4^{A_{i4}} \quad ,$$

with constraint (3.3). In the simulation procedure, we will use this joint distribution to find the posterior distribution of each parameter and simulate the parameter given the rest.

After stating the joint posterior distribution of the parameters , we may employ Gibbs sampler to impute all the parameters and A_{ij}'s, given the observed frequencies. The main theme of the method is that it exploits the simplicity of the distribution of one component of the missing values given both the observed data and the remainder of the missing values. Introduction of Gibbs sampler and the different examples of its application can be found in the articles (Tanner and Wong (1987), Li (1988) and Casella and George (1992)). We will discuss the simulation of each component in the model separately.

3.1.2 Simulation method

The simulation of each component will be presented here. Detailed methods will be discussed.

3.1.2.1 Simulation of $A_{i1},A_{i2},A_{i3},A_{i4}$

The conditional posterior distribution of $A_{i1},A_{i2},A_{i3},A_{i4}$ is proportional to

$$\binom{m}{A_{i1}, A_{i2}, A_{i3}, A_{i4}} P_1^{A_{i1}} P_2^{A_{i2}} P_3^{A_{i3}} P_4^{A_{i4}}, \qquad i = 1,\cdots,N,$$

which is multinomial(m,P_1,P_2,P_3,P_4), with constraint (3.3). Therefore, we can simply generate A_{i1}, A_{i2}, A_{i3} and A_{i4} using acceptance-rejection method:-

(a) Generate A_{i1}, A_{i2}, A_{i3} and A_{i4} from multinomial(m,P_1,P_2,P_3,P_4)

(b) If constraint (3.3) is not satisfied, goto (a); otherwise accept A_{i1}, A_{i2}, A_{i3} and A_{i4}.

However, in practice, it was found that the rejection rate was very high. Therefore, a modified method (making use of Gibbs sampler) is used to simulate A_{i1}, A_{i2}, A_{i3} and A_{i4} as follows:-

(a) given A_{i1} and A_{i2}, simulate A^*_{i3} from $B(m-A_{i1}-A_{i2},P_3/(P_3+P_4))$ with constraint $r_{2z_{i2}-1} \le A_{i2}+A^*_{i3} < r_{2z_{i2}}$, where $B(N,p)$ is the binomial distribution with parameters N and p.

(b) given A_{i2} and A^*_{i3}, simulate A^*_{i1} from $B(m-A_{i2}-A^*_{i3},P_1/(P_1+P_4))$ with constraint $r_{1z_{i1}-1} \le A^*_{i1}+A_{i2} < r_{1z_{i1}}$. A^*_{i4} is given by $m-A_{i2}-A^*_{i3}-A^*_{i1}$.

(c) given A^*_{i3} and A^*_{i4}, simulate A^*_{i2} from $B(m-A^*_{i3}-A^*_{i4},P_2/(P_1+P_2))$ with constraint $r_{2z_{i2}-1} \le A^*_{i2}+A^*_{i3} < r_{2z_{i2}}$. Update A^*_{i1} by $m-A^*_{i2}-A^*_{i3}-A^*_{i4}$.

(d) We thus get the simulated A^*_{i1}, A^*_{i2}, A^*_{i3} and A^*_{i4} which satisfy the constraint (3.3).

In the above procedure, steps (a), (b) and (c) are required to simulate binomial variates under restrictions. The following procedure is used to demonstrate how to simulate X from B(N,p) with constraint $a \le X \le b$ by a more efficient method, where a and b are integers and $0 \le a \le b \le N$.

(a) If a=b, then X=a and stop.

(b) If a=0 and b=N, simulate X from B(N,p) and stop.

(c) If N*p<a, goto (f).

(d) Simulate $Y_{b+1:N}$ which is the (b+1)th order statistics of U(0,1) of size m under the restriction $Y_{b+1:N} > p$:

   (1) Find $\pi = Pr(B \le p)$ where $B \sim Beta(a,N-a+1)$.

   (2) Simulate $U \sim U(0,1)$.

   (3) Then $Y_{b+1:N}$ is the solution of the equation $\pi+(1-\pi)*U = Pr(B \le Y_{b+1:N})$.

(e) Simulate Z, where $Z \sim Binomial(b,p/Y_{b+1:N})$. If $Z \ge a$ then X=Z and stop; otherwise goto (d).

(f) Simulate $Y_{a:N} < p$:

   (1) Find $\pi = Pr(B \le p)$ where $B \sim Beta(a,N-a+1)$.

   (2) Simulate $U \sim U(0,1)$.

   (3) Then $Y_{a:N}$ is the solution of the equation $p*U = Pr(B \le Y_{a:N})$ where $B \sim Beta(a,N-a+1)$.

(g) Simulate Z, where $Z \sim Binomial\left(N-a,(p-Y_{a:N})/(1-Y_{a:N})\right)$. If $Z+a \le b$ then X=Z and stop; otherwise goto (f).

Remarks:

(i) If $X \sim U(0,1)$; then $X_{i:n} \sim Beta\ (i,n-i+1)$. This result justify the steps (d) and (f).

(ii) Algorithm for computing the incomplete beta integral is given by Majumder and Bhattacharjee (1973).

3.1.2.2 Simulation of T

Given A_{ij}, $j = 1,\cdots,4$; $i = 1,\cdots, N$, ρ, $r_{v,1},\ldots,r_{v,n(v)-1}$, $v = 1,2$ and the observed data, the posterior distribution of T is proportional to

$$\prod_{i=1}^{N} P_1^{Ai1} P_2^{Ai2} P_3^{Ai3} P_4^{Ai4} = P_1^{\Sigma Ai1} P_2^{\Sigma Ai2} P_3^{\Sigma Ai3} P_4^{\Sigma Ai4}$$

with constraint (3.3). Therefore, the simulation procedure of T is as follows:-

(a) We simulate T from the uniform rectangle given by four vertices (V_1,V_2), (V_1,V_3), (V_4,V_3) and (V_4,V_2) where

$V_1 = \mathrm{Min}\Big\{$1st coordinate of points which are greater than T_1 in the union of the reference sets R_i's.$\Big\}$

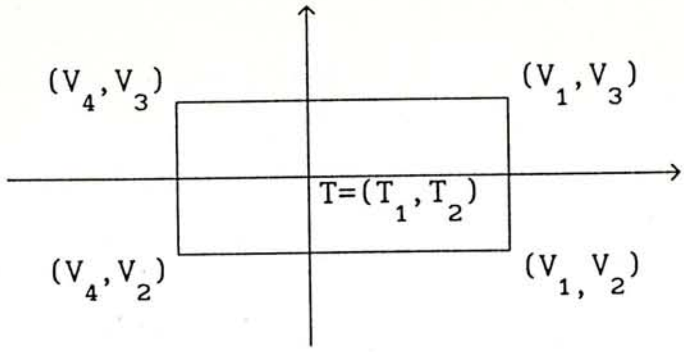
$V_2 = \mathrm{Min}\Big\{$2nd coordinate of points which are smaller than T_2 in the union of the reference sets R_i's.$\Big\}$

$V_3 = \mathrm{Max}\Big\{$2nd coordinate of points which are greater than T_2 in the union of the reference sets R_i's.$\Big\}$

$V_4 = \mathrm{Max}\Big\{$1st coordinate of points which are smaller than T_1 in the union of the reference sets R_i's.$\Big\}$

(The method for simulating V_j , $j=1,\cdots,4$, is given after the simulation procedure of T.)

i.e.



Using this method, we can ensure that the $\Sigma A_{i1}, \Sigma A_{i2}, \Sigma A_{i3}$ and ΣA_{i4} will be the same after the simulation of T as they are given in advance. (Note that A_{ij}'s depend on T.)

(b) We have to find an envelope for the acceptance-rejection method in simulating T. Since T is bounded within the rectangle, instead of using the usual envelope, $\left(\hat{P}_1\right)^{\Sigma A_{i1}} \left(\hat{P}_2\right)^{\Sigma A_{i2}} \left(\hat{P}_3\right)^{\Sigma A_{i3}} \left(\hat{P}_4\right)^{\Sigma A_{i4}}$, where $\hat{P}_j = \frac{\sum_i A_{ij}}{mN}$, $j = 1, \cdots, 4$. We find a better envelope by following method:

(i) Define $P_{Min}(j) = \underset{\substack{T \in \text{the} \\ \text{rectangle}}}{\mathrm{Min}} P_j$ and $P_{Max}(j) = \underset{\substack{T \in \text{the} \\ \text{rectangle}}}{\mathrm{Max}} P_j$, for $j = 1, \cdots, 4$.

(ii) Set $\hat{P}_j = \frac{\sum_i A_{ij}}{mN}$, for $j = 1, \cdots, 4$.

(iii) Re-calculate $\hat{P}_1$ such that

$$\hat{P}_1 = \begin{cases} P_{Min}(1) & \text{if} \quad P_{Min}(1) \geq C \\ C & \text{if} \quad P_{Min}(1) < C < P_{Max}(1) \\ P_{Max}(1) & \text{if} \quad C \geq P_{Max}(1) \end{cases},$$

$$\text{where } C = \frac{\frac{\sum_i A_{i1}}{mN} * (1 - \hat{P}_2 - \hat{P}_3)}{\frac{\sum_i A_{i1}}{mN} + \frac{\sum_i A_{i4}}{mN}}.$$

(iv) Re-calculate $\hat{P}_2$ such that

$$\hat{P}_2 = \begin{cases} P_{Min}(2) & \text{if} \quad P_{Min}(2) \geq C \\ C & \text{if} \quad P_{Min}(2) < C < P_{Max}(2) \\ P_{Max}(2) & \text{if} \quad C \geq P_{Max}(2) \end{cases},$$

$$\text{where } C = \frac{\frac{\sum_i A_{i2}}{mN} * (1 - \hat{P}_1 - \hat{P}_3)}{\frac{\sum_i A_{i2}}{mN} + \frac{\sum_i A_{i4}}{mN}}.$$

(v) Re-calculate $\hat{P}_3$ such that

$$\hat{P}_3 = \begin{cases} P_{Min}(3) & \text{if} \quad P_{Min}(3) \geq C \\ C & \text{if} \quad P_{Min}(3) < C < P_{Max}(3) \\ P_{Max}(3) & \text{if} \quad C \geq P_{Max}(3) \end{cases},$$

$$\text{where } C = \frac{\frac{\sum_i A_{i3}}{mN} * (1 - \hat{P}_2 - \hat{P}_3)}{\frac{\sum_i A_{i3}}{mN} + \frac{\sum_i A_{i4}}{mN}}.$$

(vi) Re-calculate $\hat{P}_4$ such that $\hat{P}_4 = 1 - \sum_{j=1}^{3} \hat{P}_j$.

(vii) The new envelope envelope is given by

$$\left(\hat{P}_1\right)^{\Sigma Ai1} \left(\hat{P}_2\right)^{\Sigma Ai2} \left(\hat{P}_3\right)^{\Sigma Ai3} \left(\hat{P}_4\right)^{\Sigma Ai4} ,$$

where P_j , $j = 1,\cdots,4$, is given by (iii) to (iv).

(c) We simulate the value of T (say T*) from the uniform rectangle with vertices $(V_1,V_2),(V_1,V_3),(V_4,V_3)$ and (V_4,V_2) and calculate the probability P_i, $i=1,\cdots,4$.

(d) Simulate U from U(0,1).

(e) If $$U^* \left(\hat{P}_1\right)^{\Sigma Ai1} \left(\hat{P}_2\right)^{\Sigma Ai2} \left(\hat{P}_3\right)^{\Sigma Ai3} \left(\hat{P}_4\right)^{\Sigma Ai4} > P_1^{\Sigma Ai1} P_2^{\Sigma Ai2} P_3^{\Sigma Ai3} P_4^{\Sigma Ai4}$$

then reject T^* and goto (c); otherwise accept T^*.

We now demonstrate the method of simulating the vertices V_j , $j=1,\cdots,4$ of the rectangle. In the following, only the procedure for simulating V_1 is shown as the method is similar for the rest. The simulation of the vertex V_1 is as follows:-

(a) First of all, we note that $V_1=\text{Min}\ \{X_{13}, X_{14}\}$ where X_{13} and X_{14} are the values such that

$$X_{13} = \text{Min} \left\{ Y_{ij1} \;\middle|\; (Y_{ij1}, Y_{ij2})' \in \text{Region } 3 \right\},$$

and

$$X_{14} = \text{Min} \left\{ Y_{ij1} \;\middle|\; (Y_{ij1}, Y_{ij2})' \in \text{Region } 4 \right\}.$$

(b) Secondly, the distribution of Y_{ij1} such that $(Y_{ij1}, Y_{ij2})' \in$ Region 3 is given by

$$F(X_{13}) = \frac{P_c}{P_c + P_D},$$

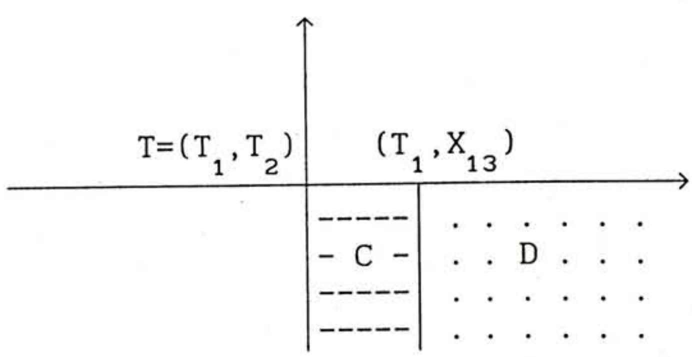
where

$$P_c = \Pr\left((Y_{ij1}, Y_{ij2})' \in \textbf{region C} \right),$$

and

$$P_D = \Pr\left((Y_{ij1}, Y_{ij2})' \in \textbf{region D} \right).$$

and regions C and D are given in the following graph:-



In order to simulate the minimum value of those values, we simulate the order statistics $U_{1:m}$ which is the first order statistic of a random sample of size m from U(0,1) and solve the equation

$$\frac{P_c}{P_c + P_D} = U_{1:m} .$$

Let the value be X_{13}. (Note that P_c and P_D are the functions of X_{13}.)

(c) By similar argument, we can find the value of X_{14}.

(d) Thus, the value of V_1 is given by V_1=Min $\{X_{13,} X_{14}\}$.

Remarks:

(i) For the formation of V_4, the method used is similar to that of V_1.

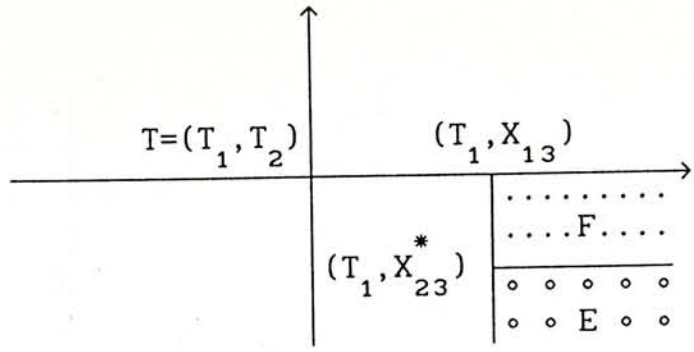
(ii) For the formation of V_2, V_3, the idea is similar to that of V_1 however, for example, X_{23} (the maximum of the values which are less than T_2 in the reference set in region 3) is found by comparing the following two values:

a) Value from the conditional distribution given X_{13} which is simulated in above.

b) Value say X^*_{23} that satisfies the following equation (the argument is similar for finding the value X_{13}),

$$\frac{P_E}{P_E + P_F} = U_{A_{i3}-1:A_{i3}-1}$$

where $U_{A_{i3}-1:A_{i3}-1}$ is the $A_{i3}-1$ th order statistic of a random sample of size $A_{i3}-1$ from U(0,1) and P_E and P_F, which have the similar definition of P_c and P_D mentioned above, are the probabilities corresponding to the regions E and F are given in the following figure:-



The value of X_{23} is thus given by the maximum of the value in a) and b).

3.1.2.3 Simulation of $r_{1,1},\ldots,r_{1,n(1)-1}$ and $r_{2,1},\ldots,r_{2,n(2)-1}$

First of all, we define

$$\mathrm{Min}(r_{1g})=\underset{i}{\mathrm{Min}}\left\{A_{i1}+A_{i2}\,\middle|\,k_{i1}=g \quad , \; i=1,\ldots,N\right\} \quad , \; g=2,\ldots,n(1)$$

$$\mathrm{Max}(r_{1g})=\underset{i}{\mathrm{Max}}\left\{A_{i1}+A_{i2}\,\middle|\,k_{i1}=g \quad , \; i=1,\ldots,N\right\} \quad , \; g=1,\ldots,n(1)-1$$

$$\mathrm{Min}(r_{2g})=\underset{i}{\mathrm{Min}}\left\{A_{i2}+A_{i3}\,\middle|\,k_{i2}=g \quad , \; i=1,\ldots,N\right\} \quad , \; g=2,\ldots,n(2)$$

$$\mathrm{Max}(r_{2g})=\underset{i}{\mathrm{Max}}\left\{A_{i2}+A_{i3}\,\middle|\,k_{i2}=g \quad , \; i=1,\ldots,N\right\} \quad , \; g=1,\ldots,n(2)-1$$

Then, we can simulate $r_{1,1},\ldots,r_{1,n(1)-1}$ and $r_{2,1},\ldots,r_{2,n(2)-1}$ using the following procedure:

(a) Simulate r_{1g} from discrete uniform $(\mathrm{Max}(r_{1g}),\mathrm{Min}(r_{1g+1}))$, $g=1,\ldots,n(1)-1$.

(b) Simulate r_{2g} from discrete uniform $(\mathrm{Max}(r_{2g}),\mathrm{Min}(r_{2g+1}))$, $g=1,\ldots,n(2)-1$.

Remarks:

Suppose no observation exists in a whole column or row like the following two contingency tables:

(i)	d1	d2	d3
c1	2	3	5
c2	3	4	6
c3	0	0	0
c4	6	4	3

(ii)	d1	d2	d3
c1	2	3	0
c2	3	4	0
c3	6	4	0
c4	6	4	0

For example, in case (i), $Max(r_{23})$ and $Min(r_{23})$ cannot be obtained as the third row has no observation. Thus, we will simulate $r_{23} = \lceil r^*_{23} \rceil$ where r^*_{23} is the largest order statistics (order number is the number of thresholds involved, in case (i), that is 2) of $U(Max(r_{22}),m)$ where m is the size of the reference set and $r_{22} = \lceil r^*_{22} \rceil$ where r^*_{22} is from $U(Max(r_{22}),r_{23})$.

3.1.2.4 Simulation of ρ

The conditional posterior distribution of ρ is proportional to

$$\prod_{i=1}^{N} P_1^{Ai1} P_2^{Ai2} P_3^{Ai3} P_4^{Ai4} .$$

We can also employ the acceptance-rejection method to simulate ρ using the following steps:-

(a) Choose an envelope for the acceptance-rejection method to be

$$\left(\hat{P}_1\right)^{\Sigma Ai1} \left(\hat{P}_2\right)^{\Sigma Ai2} \left(\hat{P}_3\right)^{\Sigma Ai3} \left(\hat{P}_4\right)^{\Sigma Ai4},$$

where $\hat{P}_j = \frac{\sum_i A_{ij}}{mN}$, $j = 1,\cdots,4$.

(b) Simulate ρ^* from $U(-1,1)$ and calculate the probability P_j, $j=1,\cdots,4$.

(c) Simulate U from U(0,1)

(d) If $$U^*\left(\hat{P}_1\right)^{\Sigma A_{i1}}\left(\hat{P}_2\right)^{\Sigma A_{i2}}\left(\hat{P}_3\right)^{\Sigma A_{i3}}\left(\hat{P}_4\right)^{\Sigma A_{i4}} > P_1^{\Sigma A_{i1}} P_2^{\Sigma A_{i2}} P_3^{\Sigma A_{i3}} P_4^{\Sigma A_{i4}}$$

then reject ρ^* and goto (b); otherwise accept ρ^*.

Remarks:

(i) A simple estimate of P_j is given by $\frac{\Sigma A_{ij}}{mN}$, $j=1,\cdots,4$; however, in practice, it is found that the rejection rate for using this bound was very high. A better bound can be found by solving the following systems of equation:- (In the following systems of equation, the symbol '*' represents the multiplier operator.)

$$\begin{cases} \hat{P}_1 + \hat{P}_2 = c1 \quad \text{where } c1 = \Phi(T_1) \\ \hat{P}_2 + \hat{P}_3 = c2 \quad \text{where } c2 = \Phi(T_2) \\ \hat{P}_1 + \hat{P}_2 + \hat{P}_3 + \hat{P}_4 = 1 \\ \frac{\partial F}{\partial \hat{P}_1} = 0 \quad \text{where } F = \sum_{j=1}^{4} X_j * \ln(\hat{P}(j)) \text{ where } X_j = \sum_i A_{ij} \end{cases}$$

which is equivalent to

$$\begin{cases} \hat{P}_2 = c1 - \hat{P}_1 \\ \hat{P}_3 = c2 - \hat{P}_2 = c2 - c1 + \hat{P}_1 \\ \hat{P}_4 = 1 - \hat{P}_1 - \hat{P}_2 - \hat{P}_3 = 1 - c2 - \hat{P}_1 \\ \frac{X_1}{\hat{P}_1} - \frac{X_2}{c1 - \hat{P}_1} + \frac{X_3}{c2 - c1 + \hat{P}_1} - \frac{X_4}{1 - c2 - \hat{P}_1} = 0 \end{cases}$$

3.1.2.5 The aggregate simulation procedure

(a) Denote the initial guess for the parameters by $\rho^{(0)}, T^{(0)}, r^{(0)}_{1,1}, \ldots, r^{(0)}_{1,n(1)-1}$ and $r^{(0)}_{2,1}, \ldots, r^{(0)}_{2,n(2)-1}$. (The methods of finding initial guess of the parameters are given in the following section 3.1.2.6.)

(b) Set k to 1.

For the k th iteration,

(c) Given the rest of parameters in iteration k-1 and the observed frequencies, simulate $A^{(k)}_{i1}, A^{(k)}_{i2}, A^{(k)}_{i3}, A^{(k)}_{i4}$ by the method given in section 3.1.2.1.

(d) Given $A^{(k)}_{i1}, A^{(k)}_{i2}, A^{(k)}_{i3}, A^{(k)}_{i4}$ obtained in (c), the rest of parameters in iteration k-1 and the observed frequencies, simulate $T^{(k)}$ by the method given in section 3.1.2.2.

(e) Given $A^{(k)}_{i1}, A^{(k)}_{i2}, A^{(k)}_{i3}, A^{(k)}_{i4}$ and $T^{(k)}$ obtained in (c) and (d) respectively, the rest of parameters in iteration k-1 and the observed frequencies, simulate $r^{(k)}_{11}, \ldots, r^{(k)}_{1n(1)-1}$ and $r^{(k)}_{21}, \ldots, r^{(k)}_{2n(2)-1}$ by the method given in section 3.1.2.3.

(f) Given the rest of parameters in iteration k and the observed frequencies, simulate $\rho^{(k)}$ by the method given in section 3.1.2.4.

(g) Increase k by 1 and goto (c) if k does not reach a pre-assigned constant or further improvement is required.

3.1.2.6 Initial guesses of the parameters

(a) Initial guess of ρ : Since we generate more than one complete data set, it is recommended to use different initial guesses of ρ in different imputation.

(b) Initial guess of $r_{1,1},\cdots,r_{1,n(1)-1}$ and $r_{2,1},\cdots,r_{2,n(2)-1}$: the initial guess of the order numbers of the thresholds are given by a simple rule. As discussed above, we can see that the thresholds will converge to the population percentile when the size of the reference set, m, tends to infinity. Moreover, when we classify the subjects, it is natural to make use of the principle of average, i.e. when we classify between two categories, we use the 50 percentile to classify; when we classify between three categories, we use 33 percentile and 66 percentile and so on. Therefore, the initial guess of $r_{1,1},\cdots,r_{1,n(1)-1}$ and $r_{2,1},\cdots,r_{2,n(2)-1}$ will be given by

$$r_{v,e} = \left\lfloor \frac{em}{n(v)} \right\rfloor , \quad e = 1,\cdots,n(v)-1; \ v=1,2 ,$$

where $\lfloor X \rfloor$ is the largest integer smaller or equal to X.

(c) Initial guess of T : Given the initial guess of $r_{1,1},\cdots,r_{1,n(1)-1}$ and $r_{2,1},\cdots,r_{2,n(2)-1}$, we find the cumulative marginal proportion of the table which is the nearest to 0.5, say $\pi(v)$, v=1,2. Moreover, we denote the number of categories included in the cumulative marginal proportion of the table to be $\eta(v)$, v=1,2. i.e.

$$\sum_{e=1}^{\eta(1)} p_{e.} = \pi(1) ,$$

where $p_{e.} = \sum_{j=1}^{n(2)} p_{ej}$ and p_{ej} is the empirical probability of the cell (e,j) and

$$\sum_{e=1}^{\eta(2)} p_{.e} = \pi(2) ,$$

where $p_{.e} = \sum_{i=1}^{n(1)} p_{ie}$ and p_{ie} is the empirical probability of the cell (i,e). Then we can have the initial guess of T_v, v=1,2, given by

$$T_v = \Phi^{-1}(Y_{r_{v,\eta(v)}:m}) , \quad v=1,2.$$

3.2 Simulation result

3.2.1 Setup

In this section, we will present the result of the simulation on the model described in Example 2.3.4.1. We select the true values of the parameters as follows:-

(1) Sample size, N = 500,

(2) n(1) = 3; n(2) =4,

(3) The size of reference set, m = 20 (Note that we cannot take m to be a very large number as the present situation is the case of 'observe without error'; otherwise, the response of the individuals will have high probability to be the same which is not good for demonstrating the worthiness of the proposed model.)

(4) The order number of the thresholds :

For variable 1, $r_{1,1}$=9; $r_{1,2}$=12.

For variable 2, $r_{2,1}$=7; $r_{2,2}$=11; $r_{2,3}$=14.

(5) The correlation : By symmetry, only positive correlation cases are considered in the simulation. Six correlations, (a) 0.95, (b) 0.75, (c) 0.55, (d) 0.35, (e) 0.15, and (f) 0.05

are considered. It covers the range of the positive correlations so that the simulation results can represent most cases in practice. The following table lists the values of ρ and their corresponding τ's and ω's.

Table 3.2.1.1 Values of ρ, τ and ω in different cases

	ρ	τ	ω
(a)	0.95000	0.79783	0.94531
(b)	0.75000	0.53989	0.73414
(c)	0.55000	0.37074	0.53207
(d)	0.35000	0.22764	0.33596
(e)	0.15000	0.09585	0.14337
(f)	0.05000	0.03148	0.04775

(6) True value of T = (0.00,0.00).

(7) Based on the selected sets of parameters, we simulated the following six contingency tables.

Table 3.2.1.1 Contingency table of the six cases simulated

(a)

24	92	9	0
3	159	101	3
0	10	78	21

(b)

21	80	22	2
14	137	104	11
1	31	59	18

(c)

19	74	28	4
16	141	92	17
2	43	46	18

(d)

17	69	35	4
12	144	96	14
4	46	44	15

(e)

11	68	39	7
19	136	93	18
5	51	43	10

(f)

7	64	47	7
21	128	99	18
6	57	37	9

(8) For each case, ten complete data sets will be generated.

3.2.2 Results

The estimates of the parameters, τ, ω, r_{ij}'s and T will be given in the following tables. For the cases of r_{ij}'s and T, sample mean and standard error of the estimate will be calculated from the estimates of the ten imputed data sets. As τ and ω are our main interest, some efforts are made to estimate their posterior variances. In the last iteration, not only one τ (similar situation for the case of ω) was generated for each complete data set. Instead 10 samples of τ were generated. As a result, we got 100 estimates of τ finally, Let us denote them by $\hat{\tau}_{ij}$, i=1,$\cdots$,10, (10 independent complete data set were simulated as mentioned in previous section.), j=1,$\cdots$,10, (number of replicates in each complete data set). Making use of these 100 estimates, we can have

(a) The mean estimate of τ for the i th complete data set is

$$\hat{\tau}_{i.} = \frac{\sum_{j=1}^{10} \hat{\tau}_{ij}}{10}, \quad i=1,\cdots,10.$$

(b) The final estimate of τ is $\hat{\tau} = \dfrac{\sum_{i=1}^{10} \hat{\tau}_{i.}}{10}$. It is the estimate of the posterior mean.

(c) The average of variance estimate is given by

$$\frac{1}{10}\sum_{i=1}^{10}\left(\frac{\sum_{j=1}^{10}(\hat{\tau}_{ij}-\hat{\tau}_{i.})^2}{9}\right).$$

It is the estimate of E(Var(τ|imputed data)|observed data).

(d) The variance estimate of the mean estimate $\hat{\tau}_{i.}$ is given by

$$\frac{\sum_{i=1}^{10}(\hat{\tau}_{i.}-\hat{\tau})^2}{9}.$$

It estimates

Var(E(τ|imputed data)|observed data)

+ E(Var(τ|imputed data)|observed data)/10

(e) The estimate of the posterior variance is given by the value in 0.9 of (c) plus the value in (d).

(f) The variability due to the incompleteness of data given by

$$\frac{\text{Value in (d) - 1/10 of value in (c)}}{\text{Posterior variance estimate}} \times 100\%$$

Following two tables summarize the estimation of τ and ω. The values discussed above are also given.

Table 3.2.2.1 Estimation of τ (Number of iterations =100)

	(a)	(b)	(c)
True value	0.79783	0.53989	0.37074
Posterior mean estimate	0.87103	0.64401	0.41602
Mean of sample Var	0.00003	0.00005	0.00008
Var of mean estimate	0.00015	0.02666	0.00683
Posterior Var estimate	0.00018	0.02671	0.00690
Posterior SD estimate	0.01330	0.16342	0.08308
% of variability due to incompleteness of data	81.67%	99.79%	98.87%

	(d)	(e)	(f)
True value	0.22764	0.09585	0.03184
Posterior mean estimate	0.18069	-0.13157	-0.21383
Mean of sample Var	0.00011	0.00009	0.00008
Var of mean estimate	0.00196	0.02245	0.01957
Posterior Var estimate	0.00206	0.02253	0.01964
Posterior SD estimate	0.04538	0.15010	0.14015
% of variability due to incompleteness of data	94.61%	99.60%	99.60%

Table 3.2.2.2 Estimation of ω (Number of iterations=100)

	(a)	(b)	(c)
True value	0.94531	0.73414	0.53207
Posterior mean estimate	0.97732	0.81315	0.58604
Mean of sample Var	0.00000	0.00004	0.00012
Var of mean estimate	0.00002	0.02468	0.01060
Posterior Var estimate	0.00002	0.02472	0.01071
Posterior SD estimate	0.00441	0.15721	0.10348
% of variability due to incompleteness of data	83.29%	99.82%	98.86%

	(d)	(e)	(f)
True value	0.33596	0.14337	0.04775
Posterior mean estimate	0.26781	-0.19213	-0.30944
Mean of sample Var	0.00023	0.00019	0.00016
Var of mean estimate	0.00413	0.04722	0.03644
Posterior Var estimate	0.00434	0.04739	0.03658
Posterior SD estimate	0.06586	0.21769	0.19127
% of variability due to incompleteness of data	94.63%	99.60%	99.57%

In tables 3.2.2.3 and 3.2.2.4, we list the estimates of T and the order number r_{ij}'s. It should be noted that the standard error is the standard error of the estimate, which is the estimate of the posterior mean. Thus estimate ± 2 (standard error) is an approximate 95% confidence interval of the posterior mean, which can of course be different from the true parameter value.

Table 3.2.2.3 Estimation of T

	(a)	(b)	(c)
True value	(0.0, 0.0)	(0.0, 0.0)	(0.0, 0.0)
Estimate	(-0.34964, -0.11408)	(-0.36477, -0.10646)	(-0.35787, -0.11803)
Standard error	(0.00392, 0.00101)	(0.00363, 0.00110)	(0.00465, 0.00109)

	(d)	(e)	(f)
True value	(0.0, 0.0)	(0.0, 0.0)	(0.0, 0.0)
Estimate	(-0.35616, -0.11700)	(-0.35300, -0.10281)	(-0.35638, -0.09399)
Standard error	(0.00659, 0.00068)	(0.00526, 0.00244)	(0.00481, 0.00315)

Table 3.2.2.4 Estimation of the order number of the thresholds

	(a)	(b)	(c)
True value	(9.0,12.0) (7.0,11.0, 14.0)	(9.0,12.0) (7.0,11.0, 14.0)	(9.0,12.0) (7.0,11.0, 14.0)
Estimate	(6.00000, 12.10000) (5.30000, 10.20000, 15.00000)	(6.00000, 11.40000) (5.30000, 10.10000, 15.00000)	(6.00000, 12.30000) (5.10000, 10.00000, 15.00000)
Standard error	(0.00000, 0.23333) (0.15275, 0.13333, 0.00000)	(0.00000, 0.37118) (0.21344, 0.10000, 0.00000)	(0.00000, 0.33500) (0.10000, 0.00000, 0.00000)

	(d)	(e)	(f)
True value	(9.0,12.0) (7.0,11.0, 14.0)	(9.0,12.0) (7.0,11.0, 14.0)	(9.0,12.0) (7.0,11.0, 14.0)
Estimate	(6.00000, 12.10000) (5.30000, 10.00000, 14.90000)	(5.90000, 11.80000) (5.30000, 10.10000, 15.00000)	(6.00000, 12.00000) (5.40000, 10.00000, 15.00000)
Standard error	(0.00000, 0.17275) (0.15275, 0.00000, 0.10000)	(0.10000, 0.38873) (0.15275, 0.00000, 0.00000)	(0.00000, 0.44721) (0.16330, 0.00000, 0.00000)

3.2.3 Discussion

(1) First of all, the reason for choosing the number of iterations to be 100 is that when trying greater numbers of iterations, like 150, 200 and 250, it was found that there is no significant improvement in both the estimate and the standard error of the estimate. It can be assumed that after 100 iterations, the distribution of the estimators has converged to their posterior distribution. On the other hand, it was found that the required number of iterations for convergence depends on the values of the parameters. Therefore, the stopping rule of 100 iterations is not a fixed rule for the convergence in general.

(2) The estimates for τ and ω distribution shows that the

estimation is quite well in the cases of (b) to (f). The true value lies within two posterior standard deviation of the posterior mean. In case (a), i.e. ρ=0.95, the imprecise of the estimation may be owing to two reasons. Firstly, we believe that the posterior distribution of ρ is not of normal shape and should be skew to the left as 0.95 is close to the boundary of the space of ρ. Secondly, the computational errors in the calculation of the bivariate normal distribution function will have a larger effect on the estimates.

(3) As we have not estimated the posterior variances of T's and the order numbers, r_{ij}'s, it is hard to comment on the performance of the estimates.

(4) Some estimates of the order number remain constant throughout the whole procedure. The problem may be due to (i) the small values of the size of reference set m, and (ii) the discreteness of the order numbers.

Chapter 4 Discussion

In the present article, we have discussed various aspects of the vaguely classified variables: the meanings of vaguely classified variables; the model on the classification of subjects into the categories of the vaguely classified variables by making use of reference set; the existence of super-population; the relationship of the super population and the reference set; and the effect of 'observing with error' on the classification.

A model is proposed to describe the classification of the vaguely classified variables in an elegant and precise manner. Moreover, by making use of a special case of the proposed model, we illustrate the estimation method of the interesting parameters. From the simulation study discussed in the previous chapter, it is found that the method used performs quite well in estimation in most of the cases. Furthermore, it is believed that for the other cases of the proposed model, like the dimension of the variables is greater than two or the situation of observing with error, this approach can also handle in ease. However, when the model become more complicated, an efficient implementation of the simulation is the key to success. Like the present situation of simulating A_{ij} , $j=1,\cdots,4$ and $i=1,\cdots,N$, originally, we have to simulate A_{ij}, $j=1,\cdots,4$ from multinomial (m,P_1,P_2,P_3,P_4) with constraint (3.3). However, owing to the high rejection rate in the acceptance rejection method, we have to change the procedure to simulate two of A_{ij}'s given the rest. Therefore, it is foreseen that a large amount of similar simulation problems will be aroused when dealing with more general models. As a result, a large amount of efforts is required in order to achieve the aim.

Secondly, the program used for imputation is not efficient enough in the sense that like the example in the previous paragraph, we have to simulate two of A_{ij}'s given the rest, instead of simulating A_{ij}'s simultaneously. As a result, the CPU time required for completing an iteration will be longer. Further investigations are needed to improve some simulation steps.

Thirdly, the proposed model is a generalization of the fixed-thresholds model because when m(i) (the size of the reference set) tends to infinity, the reference set of each individual i will be the same as the super-population. As we require the order number of the thresholds $\frac{r_{v,e}}{m(i)}$ converges in distribution to a fixed constant which is independent of i, say $\phi(v,e)$, $v=1,\cdots,p$; $e=1,\cdots,n(v)-1$; $i=1,\cdots,N$. The threshold $\xi_{r_{v,e}}$ will converge in probability to the percentile of the super-population, i.e.

$$\xi_{r_{v,e}} \overset{P}{\to} \varphi_{\phi(v,e)} .$$

The percentile $\varphi_{\phi(v,e)}$ is an unknown but fixed constant. As a result, test of adequacy of the fixed threshold model will be possible as the fixed thresholds model is a special case of the present model.

The concept of fuzzy sets closely relates to the vaguely classified variables. There is no doubt that the vaguely classified variables have the concept of fuzziness. 'How tall is tall?', 'How good is good?' are typical questions discussed in the context of fuzziness as well as that of the vaguely classified variables. The proposed model actually is an intention to give a reasonable and convincing way to describe what are behind those mystery questions. It is made as close to the reality as possible by making use of random thresholds. However, the relationship of the model here and the contents of fuzzy sets, like the membership function, has not been investigated in the present article. Further study is needed to find the connection between fuzzy sets and the proposed models as they both are dealing with same kind of concept.

Finally, we have demonstrated the estimation procedure of the proposed model only. There remain a lot of aspects that have to be handled in order to complete the whole story. For example, 'Is random thresholds with constant order number model sufficient for describing the story, or is there necessity to develop a more general model', ' Besides point estimation, how to handle the other inferential problems like interval estimation and hypothesis testing?'. In brief, further investigations are needed in order to give answers to the above questions.

Reference

Benjafield J., Frommhold K., Keenan T., Muckenheim R. and Mueller D. (1993), Imagery, Concreteness, Goodness, and Familiarity Ratings for 500 Proverbs Sampled from the Oxford Dictionary of English Proverbs, *Behavior Research Methods, Instruments, & Computers*, **25**, 27-40.

Casella G. and George E. I. (1992), Explaining the Gibbs Sampler, *American Statistician*, **46**, 167-174.

Chow K. S. (1993), *Interpretation and Estimation of Membership Functions*, M. Phil. Thesis, The Chinese University of Hong Kong.

Cramér, H. (1946), *Mathematical Methods of Statistics*, Princeton, NJ:Princeton University Press, 290.

DeSarbo W. S. and Cho J. (1989), A Stochastic Multidemsional Scaling Vector Threshold Model for the Spatial Representation of "Pick Any/N" Data, *Psychometrika*, **54**, 105-129.

Farebrother R.W. (1989), Bivariate Normal Orthant Probability, *The American Statistician*, **43**, 291-292.

Giles, R. (1982), Semantics for Fuzzy Reasoning. *Int. J. Man.-Machine Studies*, **17**, 401-415.

Giles, R. (1988), The Concept of Grade of Membership, *Fuzzy set and System*, **25**, 297-323.

Helmes E. and Jackson D.N. (1989), Prediction Models of Personality Item Responding, *Multivariate Behavioral Research*, **24**, 71-91.

Hisdal, E. (1986a), Infinite-valued Logic based on Two-valued Logic and Probability. Part 1.1. Difficulties with Present-day Fuzzy-set Theory and their Resolution in the TEE model. *Int. J. Man-Machine Studies*, **25**, 89-111.

Hisdal, E. (1986b), Infinite-valued Logic based on Two-valued Logic and Probability. Part 1.2. Different Sources of Fuzziness. *Int. J. Man-Machine Studies*, **25**, 113-138.

Hisdal, E. (1988a), Are Grades of Membership Probabilities? *Fuzzy set and Systems*, **25**, 325-348.

Hisdal, E. (1988b), The Philosophical Issues Raised by Fuzzy Set Theory,*Fuzzy set and Systems*, **25**, 349-356.

Kepner J.L., Harper J.D. and Keith S.Z. (1989), A Note on Evaluating a Certain Orthant Probability, *The American Statistician*, **43**, 48-49.

Leigh J.H. and Martin C.R. JR. (1987), "Don't Know" Item Nonresponse in a Telephone Survey: Effects of Question Form and Respondent Characteristics, *Journal of Marketing Research*, **24**, 418-424.

Li K. H. (1988), Imputatin Using Markov Chains, *J. Statist. Comput. Simul.*, **30**, 57-79.

Mabuchi S. (1992), An Intrepretation of Membership Functions and the Properties of the General Probabilistic Operators as Fuzzy Set Operators - Part I: Case of Type I Fuzzy Sets, *Fuzzy Set and System*, **49**, 271-283.

Majumder K. L. and Bhattacharjee G. P. (1973), Algorithm AS 63, The Incomplete Beta Intergral, *Appl. Statist.*, **22**, 409-411.

Norwich A. M. and Turksen I. B. (1984), A Model for the Measurement of Membership and the Consequence of its Empirical Implementation. *Fuzzy Set and System*, **12**, 1-25.

Olsson U. (1979), Maximum Likelihood Estimation of the Polychoric Correlation Coefficient. *Psychometrika*, **44**, 443-460.

Orlovski, S. A. (1990), Decomposition of Properties and Definition of Fuzzy Set. *Fuzzy set and System*, **37**, 53-63.

Osherson D. N. and Smith E. E. (1981), On the Adequacy of Prototype Theory as a Theory of Concepts. *Cognition*, **9**, 35-38.

Osherson D. N. and Smith E. E. (1982), Gradedness and Conceptual Combination. *Cognition*, **12**, 299-318.

Smithson M. (1986), *Fuzzy Set Analysis for Behavioral and Social Sciences*, Springer-Verlag, New York.

Stigler S.M. (1989), Normal Orthant Probability, *The American Statistician*, **43**, 291-292.

Tanner M. A. and Wong W. H. (1987), The Calculation of Posterior Distribution by Data Augmentation, *J.A.S.A.*, **82**, 528-550.

Tong H. (1990), *Non-Linear Time Series - A Dynamical System Approach*, Oxford Science Publications.

Turksen I. B. (1991), Measurement of Membership Functions and Their Acquisition, *Fuzzy Set and System*, **40**, 5-38.

Zadeh, L.A. (1965), Fuzzy Sets. *Information and Control*, **8**, 338-353.

CUHK Libraries
000275950